

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Zavada, Jan
Pastorekova, Silvia
Pastorek, Jaromir

(ii) TITLE OF INVENTION: MN Gene and Protein

(iii) NUMBER OF SEQUENCES: 86

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Leona L. Lauder
(B) STREET: 369 Pine Street
(C) CITY: San Francisco
(D) STATE: California
(E) COUNTRY: USA
(F) ZIP: 94104

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 09/772,719
(B) FILING DATE: 01-30-2001
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/485,049
(B) FILING DATE: 07-JUN-1995

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Lauder, Leona L.
(B) REGISTRATION NUMBER: 30,863
(C) REFERENCE/DOCKET NUMBER: D-0021.3A-2

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 415-981-2034
(B) TELEFAX: 415-981-0332

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1522 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ACAGTCAGCC	GCATGGCTCC	CCTGTGCC	AGCCCCTGGC	TCCCTCTGTT	GATCCCGGCC	60
CCTGCTCCAG	GCCTCACTGT	GCAACTGCTG	CTGTCACTGC	TGCTTCTGAT	GCCTGTCCAT	120
CCCCAGAGGT	TGCCCCGGAT	GCAGGAGGAT	TCCCCCTTGG	GAGGAGGCTC	TTCTGGGGAA	180
GATGACCCAC	TGGGCGAGGA	GGATCTGCC	AGTGAAGAGG	ATTCACCCAG	AGAGGAGGAT	240
CCACCCGGAG	AGGAGGATCT	ACCTGGAGAG	GAGGATCTAC	CTGGAGAGGA	GGATCTACCT	300
GAAGTTAAC	CTAAATCAGA	AGAAGAGGGC	TCCCTGAAGT	TAGAGGATCT	ACCTACTGTT	360
GAGGCTCCTG	GAGATCCTCA	AGAACCCAG	AATAATGCC	ACAGGGACAA	AGAAGGGGAT	420
GACCAGAGTC	ATTGGCGCTA	TGGAGGCGAC	CCGCCCTGGC	CCCGGGTGTC	CCCAGCCTGC	480
GCGGGCCGCT	TCCAGTCCCC	GGTGGATATC	CGCCCCCAGC	TCGCCGCCTT	CTGCCCGGCC	540
CTGCGCCCCC	TGGAACTCCT	GGGCTTCCAG	CTCCCGCCGC	TCCCAGAACT	GCGCCTGC	600
AACAATGGCC	ACAGTGTGCA	ACTGACCTG	CCTCCTGGGC	TAGAGATGGC	TCTGGGTCCC	660
GGGCAGGGAGT	ACCGGGCTCT	GCAGCTGCAT	CTGCACTGGG	GGGCTGCAGG	TCGTCCGGGC	720
TCGGAGCACA	CTGTGGAAGG	CCACCGTTTC	CCTGCCGAGA	TCCACGTGGT	TCACCTCAGC	780

ACCGCCTTG	CCAGAGTTGA	CGAGGCCTTG	GGGCGCCCGG	GAGGCCTGGC	CGTGTGGCC	840
GCCTTCTGG	AGGAGGGCCC	GGAAGAAAAC	AGTGCCTATG	AGCAGTTGCT	GTCTCGCTTG	900
GAAGAAATCG	CTGAGGAAGG	CTCAGAGACT	CAGGTCCCAG	GACTGGACAT	ATCTGCACTC	960
CTGCCCTCTG	ACTTCAGCCG	CTACTTCAA	TATGAGGGGT	CTCTGACTAC	ACCGCCCTGT	1020
GCCCAGGGTG	TCATCTGGAC	TGTGTTAAC	CAGACAGTGA	TGCTGAGTGC	TAAGCAGCTC	1080
CACACCCCTCT	CTGACACCCT	GTGGGGACCT	GGTGACTCTC	GGCTACAGCT	GAACATTCCGA	1140
GCGACGCAGC	CTTTGAATGG	GCGAGTGATT	GAGGCCTCCT	TCCCTGCTGG	AGTGGACAGC	1200
AGTCCTCGGG	CTGCTGAGCC	AGTCCAGCTG	AATTCTGCC	TGGCTGCTGG	TGACATCCTA	1260
GCCCTGGTTT	TTGGCCTCCT	TTTGCTGTC	ACCAGCGTCG	CGTTCCCTGT	GCAGATGAGA	1320
AGGCAGCACA	GAAGGGGAAC	CAAAGGGGT	GTGAGCTACC	GCCCAGCAGA	GGTAGCCGAG	1380
ACTGGAGCCT	AGAGGCTGGA	TCTTGGAGAA	TGTGAGAAGC	CAGCCAGAGG	CATCTGAGGG	1440
GGAGCCGGTA	ACTGT CCTGT	CCTGCTCATT	ATGCCACTTC	CTTTAACTG	CCAAGAAATT	1500
TTTTAAAATA	AATATTTATA	AT				1522

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 459 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- (A) DESCRIPTION: First 37 amino acids represent signal peptide, and remaining amino acids represent mature protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met	Ala	Pro	Leu	Cys	Pro	Ser	Pro	Trp	Leu	Pro	Leu	Leu	Ile	Pro	Ala
-35								-30						-25	

Pro Ala Pro Gly Leu Thr Val Gln Leu Leu Leu Ser Leu Leu Leu Leu
-20 -15 -10

Met Pro Val His Pro Gln Arg Leu Pro Arg Met Gln Glu Asp Ser Pro
-5 1 5 10

Leu Gly Gly Ser Ser Gly Glu Asp Asp Pro Leu Gly Glu Glu Asp
15 20 25

Leu Pro Ser Glu Glu Asp Ser Pro Arg Glu Glu Asp Pro Pro Gly Glu
30 35 40

Glu Asp Leu Pro Gly Glu Glu Asp Leu Pro Gly Glu Glu Asp Leu Pro
45 50 55

Glu Val Lys Pro Lys Ser Glu Glu Glu Gly Ser Leu Lys Leu Glu Asp
60 65 70 75

Leu Pro Thr Val Glu Ala Pro Gly Asp Pro Gln Glu Pro Gln Asn Asn
80 85 90

Ala His Arg Asp Lys Glu Gly Asp Asp Gln Ser His Trp Arg Tyr Gly
95 100 105

Gly Asp Pro Pro Trp Pro Arg Val Ser Pro Ala Cys Ala Gly Arg Phe
110 115 120

Gln Ser Pro Val Asp Ile Arg Pro Gln Leu Ala Ala Phe Cys Pro Ala
125 130 135

Leu Arg Pro Leu Glu Leu Leu Gly Phe Gln Leu Pro Pro Leu Pro Glu
140 145 150 155

Leu Arg Leu Arg Asn Asn Gly His Ser Val Gln Leu Thr Leu Pro Pro
160 165 170

Gly Leu Glu Met Ala Leu Gly Pro Gly Arg Glu Tyr Arg Ala Leu Gln
175 180 185

Leu His Leu His Trp Gly Ala Ala Gly Arg Pro Gly Ser Glu His Thr
190 195 200

Val Glu Gly His Arg Phe Pro Ala Glu Ile His Val Val His Leu Ser
205 210 215

Thr Ala Phe Ala Arg Val Asp Glu Ala Leu Gly Arg Pro Gly Gly Leu
220 225 230 235

Ala Val Leu Ala Ala Phe Leu Glu Glu Gly Pro Glu Glu Asn Ser Ala
240 245 250

Tyr Glu Gln Leu Leu Ser Arg Leu Glu Glu Ile Ala Glu Glu Gly Ser
255 260 265

Glu Thr Gln Val Pro Gly Leu Asp Ile Ser Ala Leu Leu Pro Ser Asp
270 275 280

Phe Ser Arg Tyr Phe Gln Tyr Glu Gly Ser Leu Thr Thr Pro Pro Cys
285 290 295

Ala Gln Gly Val Ile Trp Thr Val Phe Asn Gln Thr Val Met Leu Ser
300 305 310 315

Ala Lys Gln Leu His Thr Leu Ser Asp Thr Leu Trp Gly Pro Gly Asp
320 325 330

Ser Arg Leu Gln Leu Asn Phe Arg Ala Thr Gln Pro Leu Asn Gly Arg
335 340 345

Val Ile Glu Ala Ser Phe Pro Ala Gly Val Asp Ser Ser Pro Arg Ala
350 355 360

Ala Glu Pro Val Gln Leu Asn Ser Cys Leu Ala Ala Gly Asp Ile Leu
365 370 375

Ala Leu Val Phe Gly Leu Leu Phe Ala Val Thr Ser Val Ala Phe Leu
380 385 390 395

Val Gln Met Arg Arg Gln His Arg Arg Gly Thr Lys Gly Gly Val Ser
400 405 410

Tyr Arg Pro Ala Glu Val Ala Glu Thr Gly Ala
415 420

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CGCCCAGTGG GTCATCTTCC CCAGAAGAG

29

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGAATCCTCC TGCATCCGG

19

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10898 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGATCCTGTT GACTCGTGAC CTTACCCCCA ACCCTGTGCT CTCTGAAACA TGAGCTGTGT 60
CCACTCAGGG TTAAATGGAT TAAGGGCGGT GCAAGATGTG CTTTGTAAA CAGATGCTTG 120
AAGGCAGCAT GCTCGTTAAG AGTCATCACCA AATCCCTAAT CTCAAGTAAT CAGGGACACA 180
AACACTGCGG AAGGCCGCAG GGTCCCTCTGC CTAGGAAAAC CAGAGACCTT TGTTCACTTG 240
TTTATCTGAC CTTCCCTCCA CTATTGTCCA TGACCCTGCC AAATCCCCCT CTGTGAGAAA 300
CACCCAAGAA TTATCAATAA AAAAATAAAAT TTAAAAAAA AATACAAAAA AAAAAAAA 360
AAAAAAA GACTTACGAA TAGTTATTGA TAAATGAATA GCTATTGGTA AAGCCAAGTA 420
AATGATCATA TTCAAAACCA GACGCCATC ATCACAGCTC AAGTCTACCT GATTTGATCT 480
CTTTATCATT GTCATTCTT GGATTCACTA GATTAGTCAT CATCCTCAAA ATTCTCCCCC 540
AAGTTCTAAT TACGTTCAA ACATTTAGGG GTTACATGAA GCTTGAACCT ACTACCTTCT 600
TTGCTTTGA GCCATGAGTT GTAGGAATGA TGAGTTACA CCTTACATGC TGGGGATTAA 660
TTTAAACTTT ACCTCTAAGT CAGTTGGGT A CCTTGGCT TATTTTGTA GCTAATTTG 720
TAGTTAATGG ATGCACTGTG AATCTTGCTA TGATAGTTT CCTCCACACT TTGCCACTAG 780
GGGTAGGTAG GTACTCAGTT TTCAGTAATT GCTTACCTAA GACCCTAAGC CCTATTTCTC 840
TTGTACTGGC CTTTATCTGT AATATGGGCA TATTTAATAC AATATAATTT TTGGAGTTT 900
TTTGTGTT TGTTTGTGTTG TTTTTTGAG ACGGAGTCTT GCATCTGTCA TGCCCAGGCT 960
GGAGTAGCAG TGGTGCCATC TCGGCTCACT GCAAGCTCCA CCTCCCGAGT TCACGCCATT 1020
TTCCTGCCTC AGCCTCCCGA GTAGCTGGGA CTACAGGCGC CGGCCACCCT GCCCGGCTAA 1080

TTTTTGTAT TTTGGTAGA GACGGGTTT CACCGTGT A GCCAGAATGG TCTCGATCTC 1140
CTGACTTCGT GATCCACCGG CCTCGGCCTC CCAAAGTTCT GGGATTACAG GTGTGAGCCA 1200
CCGCACCTGG CCAATTTTT GAGTCTTTA AAGTAAAAAT ATGTCTTGT A AGCTGGTAAC 1260
TATGGTACAT TTCCTTTAT TAATGTGGTG CTGACGGTCA TATAAGTTCT TTTGAGTTG 1320
GCATGCATAT GCTACTTTT GCAGTCCTT CATTACATT TTCTCTCTTC ATTTGAAGAG 1380
CATGTTATAT CTTTAGCTT CACTTGGCTT AAAAGGTTCT CTCATTAGCC TAACACAGTG 1440
TCATTGTTGG TACCACTTGG ATCATAAGTG GAAAACAGT CAAGAAATTG CACAGTAATA 1500
CTTGTGTTA AGAGGGATGA TTCAGGTGAA TCTGACACTA AGAAACTCCC CTACCTGAGG 1560
TCTGAGATTG CTCTGACATT GCTGTATATA GGCTTTCCCT TTGACAGCCT GTGACTGCGG 1620
ACTATTTTC TTAAGCAAGA TATGCTAAAG TTTGTGAGC CTTTTCCAG AGAGAGGTCT 1680
CATATCTGCA TCAAGTGAGA ACATATAATG TCTGCATGTT TCCATATTTC AGGAATGTTT 1740
GCTTGTGTT TATGCTTTA TATAGACAGG GAAACTGTT CCTCAGTGAC CCAAAAGAGG 1800
TGGGAATTGT TATTGGATAT CATCATTGGC CCACGCTTTC TGACCTTGG AACAATTAAG 1860
GGTCATAAT CTCATTCTG TCAGAATTGG TACAAGAAAT AGCTGCTATG TTTCTTGACA 1920
TTCCACTTGG TAGGAAATAA GAATGTGAAA CTCTTCAGTT GGTGTGTGTC CCTNGTTTT 1980
TTGCAATTTC CTTCTTACTG TGTTAAAAAA AAGTATGATC TTGCTCTGAG AGGTGAGGCA 2040
TTCTTAATCA TGATCTTAA AGATCAATAA TATAATCCTT TCAAGGATTA TGTCTTTATT 2100
ATAATAAAGA TAATTGTCT TTAACAGAAT CAATAATATA ATCCCTTAAA GGATTATATC 2160
TTTGCTGGGC GCAGTGGCTC ACACCTGTAACCCAGCACT TTGGGTGGCC AAGGTGGAAG 2220
GATCAAATT GCCTACTTCT ATATTATCTT CTAAAGCAGA ATTCACTCTCT CTTCCCTCAA 2280
TATGATGATA TTGACAGGGT TTGCCCTCAC TCACTAGATT GTGAGCTCCT GCTCAGGGCA 2340
GGTAGCGTTT TTGTTTTTG TTTTGTGTTT TCTTTTGA GAGAGGGTCT TGCTCTGTCA 2400
CCCAGGCCAG AGTGAATGG TACAGTCTCA GCTCACTGCA GCCTCAACCG CCTCGGCTCA 2460

AACCATCATC CCATTCAGC CTCCTGAGTA GCTGGGACTA CAGGCACATG CCATTACACC 2520
TGGCTAATT TTTTGTATTT CTAGTAGAGA CAGGGTTGG CCATGTTGCC CGGGCTGGTC 2580
TCGAACTCCT GGACTCAAGC AATCCACCCA CCTCAGCCTC CCAAAATGAG GGACCGTGTC 2640
TTATTCATTT CCATGTCCT AGTCCATAGC CCAGTGCTGG ACCTATGGTA GTACTAAATA 2700
AATATTGTT GAATGCAATA GTAAATAGCA TTTCAGGGAG CAAGAACTAG ATTAACAAAG 2760
GTGGTAAAAG GTTTGGAGAA AAAAATAATA GTTTAATTG GCTAGAGTAT GAGGGAGAGT 2820
AGTAGGAGAC AAGATGGAAA GGTCTCTTGG GCAAGGTTT GAAGGAAGTT GGAAGTCAGA 2880
AGTACACAAT GTGCATATCG TGGCAGGCAG TGGGGAGCCA ATGAAGGCTT TTGAGCAGGA 2940
GAGTAATGTG TTGAAAAATA AATATAGGTT AACACCTATCA GAGCCCCTCT GACACATACA 3000
CTTGCTTTTC ATTCAAGCTC AAGTTGTCT CCCACATACC CATTACTTAA CTCACCCCTCG 3060
GGCTCCCTA GCAGCCTGCC CTACCTGTTT ACCTGCTTCC TGTTGGAGTC AGGGATGTAT 3120
ACATGAGCTG CTTTCCCTCT CAGCCAGAGG ACATGGGGGG CCCCAGCTCC CCTGCCTTTC 3180
CCCTTCTGTG CCTGGAGCTG GGAAGCAGGC CAGGGTTAGC TGAGGCTGGC TGGCAAGCAG 3240
CTGGGTGGTG CCAGGGAGAG CCTGCATAGT GCCAGGTGGT GCCTTGGTT CCAAGCTAGT 3300
CCATGGCCCC GATAACCTTC TGCCTGTGCA CACACCTGCC CCTCACTCCA CCCCATCCT 3360
AGCTTGGTA TGGGGGAGAG GGCACAGGGC CAGACAAACC TGTGAGACTT TGGCTCCATC 3420
TCTGCAAAAG GGCGCTCTGT GAGTCAGCCT GCTCCCTCC AGGCTTGCTC CTCCCCCACC 3480
CAGCTCTCGT TTCCAATGCA CGTACAGCCC GTACACACCG TGTGCTGGGA CACCCACAG 3540
TCAGCCGCAT GGCTCCCTG TGCCCCAGCC CCTGGCTCCC TCTGTTGATC CGGGCCCTG 3600
CTCCAGGCCT CACTGTGCAA CTGCTGCTGT CACTGCTGCT TCTGGTGCCT GTCCATCCCC 3660
AGAGGTTGCC CGGGATGCAG GAGGATTCCC CCTTGGGAGG AGGCTCTCT GGGGAAGATG 3720
ACCCACTGGG CGAGGAGGGAT CTGCCAGTG AAGAGGATTG ACCCAGAGAG GAGGATCCAC 3780
CCGGAGAGGA GGATCTACCT GGAGAGGGAGG ATCTACCTGG AGAGGAGGAT CTACCTGAAG 3840

TTAAGCCTAA ATCAGAAGAA GAGGGCTCCC TGAAGTTAGA GGATCTACCT ACTGTTGAGG 3900
CTCCTGGAGA TCCTCAAGAA CCCCAGAATA ATGCCACAG GGACAAAGAA GGTAAGTGGT 3960
CATCAATCTC CAAATCCAGG TTCCAGGAGG TTCATGACTC CCCTCCCATA CCCCCAGCCTA 4020
GGCTCTGTTC ACTCAGGGAA GGAGGGGAGA CTGTACTCCC CACAGAAGCC CTTCCAGAGG 4080
TCCCATAACCA ATATCCCCAT CCCGACTCTC GGAGGTAGAA AGGGACAGAT GTGGAGAGAA 4140
AATAAAAAGG GTGCAAAAGG AGAGAGGTGA GCTGGATGAG ATGGGAGAGA AGGGGGAGGC 4200
TGGAGAAGAG AAAGGGATGA GAACTGCAGA TGAGAGAAAA AATGTGCAGA CAGAGGAAAA 4260
AAATAGGTGG AGAAGGAGAG TCAGAGAGTT TGAGGGGAAG AGAAAAGGAA AGCTTGGGAG 4320
GTGAAGTGGG TACCAGAGAC AAGCAAGAAG AGCTGGTAGA AGTCATCTCA TCTTAGGCTA 4380
CAATGAGGAA TTGAGACCTA GGAAGAAGGG ACACAGCAGG TAGAGAACG TGGCTTCTTG 4440
ACTCCAAGC CAGGAATTG GGGAAAGGGG TTGGAGACCA TACAAGGCAG AGGGATGAGT 4500
GGGGAGAAGA AAGAAGGGAG AAAGGAAAGA TGGTGTACTC ACTCATTGG GACTCAGGAC 4560
TGAAGTGCC C ACTCACTTT TTTTTTTTT TTTTGAGAC AAACTTCAC TTTTGTGCC 4620
CAGGCTGGAG TGCAATGGCG CGATCTCGGC TCACTGCAAC CTCCACCTCC CGGGTTCAAG 4680
TGATTCTCCT GCCTCAGCCT CTAGCCAAGT AGCTGCGATT ACAGGCATGC GCCACCACGC 4740
CCGGCTAATT TTTGTATTT TAGTAGAGAC GGGGTTCCG CATGTTGGTC AGGCTGGTCT 4800
CGAACTCCTG ATCTCAGGTG ATCCAACCAC CCTGGCCTCC CAAAGTGCTG GGATTATAGG 4860
CGTGAGCCAC AGCGCCTGGC CTGAAGCAGC CACTCACTTT TACAGACCTT AAGACAATGA 4920
TTGCAAGCTG GTAGGATTGC TGTTTGGCCC ACCCAGCTGC GGTGTTGAGT TTGGGTGCGG 4980
TCTCCTGTGC TTTGCACCTG GCCCGCTTAA GGCATTTGTT ACCCGTAATG CTCCTGTAAG 5040
GCATCTGCGT TTGTGACATC GTTTTGGTCG CCAGGAAGGG ATTGGGGCTC TAAGCTTGAG 5100
CGGTTCATCC TTTTCATTTA TACAGGGGAT GACCAGAGTC ATTGGCGCTA TGGAGGTGAG 5160
ACACCCACCC GCTGCACAGA CCCAATCTGG GAACCCAGCT CTGTGGATCT CCCCTACAGC 5220

CGTCCCTGAA CACTGGTCCC GGGCGTCCCA CCCGCCGCC ACCGTCCCAC CCCCTCACCT 5280
TTTCTACCCG GGTTCCCTAA GTTCCTGACC TAGGCGTCAG ACTTCCTCAC TATACTCTCC 5340
CACCCCCAGGC GACCCGCCCT GGCCCCGGGT GTCCCCAGCC TGCGCGGGCC GCTTCCAGTC 5400
CCCGGTGGAT ATCCGCCCCC AGCTGCCCGC CTTCTGCCCG GCCCTGCGCC CCCTGGAACT 5460
CCTGGGCTTC CAGCTCCCAGC CGCTCCCAGA ACTGCGCCTG CGCAACAATG GCCACAGTGG 5520
TGAGGGGGTC TCCCCGCCGA GACTTGGGGA TGGGGCGGGG CGCAGGGAAG GGAACCGTCG 5580
CGCAGTGCCT GCCCCGGGGT TGGGCTGGCC CTACCGGGCG GGGCCGGCTC ACTTGCCTCT 5640
CCCTACGCAG TGCAACTGAC CCTGCCTCCT GGGCTAGAGA TGGCTCTGGG TCCCGGGCGG 5700
GAGTACCGGG CTCTGCAGCT GCATCTGCAC TGGGGGGCTG CAGGTCGTCC GGGCTCGGAG 5760
CACACTGTGG AAGGCCACCG TTTCCCTGCC GAGGTGAGCG CGGACTGGCC GAGAAGGGC 5820
AAAGGAGCGG GGCGGACGGG GGCCAGAGAC GTGGCCCTCT CCTACCCCTCG TGTCCTTTTC 5880
AGATCCACGT GGTTCACCTC AGCACCGCCT TTGCCAGAGT TGACGAGGCC TTGGGGCGCC 5940
CGGGAGGCCT GGCGTGTG GCCGCCTTTC TGGAGGTACC AGATCCTGGA CACCCCTAC 6000
TCCCCGCTT CCCATCCCCT GCTCCTCCCG GACTCTATCG TGGAGCCAGA GACCCCATCC 6060
CAGCAAGCTC ACTCAGGCCCT CTGGCTGACA AACTCATTCA CGCACTGTTT GTTCATTAA 6120
CACCCACTGT GAACCAGGCA CCAGCCCCA ACAAGGATTC TGAAGCTGTA GGTCCCTGCC 6180
TCTAAGGAGC CCACAGCCAG TGGGGGAGGC TGACATGACA GACACATAGG AAGGACATAG 6240
TAAAGATGGT GGTCACAGAG GAGGTGACAC TTAAAGCCTT CACTGGTAGA AAAGAAAAGG 6300
AGGTGTTCAT TGCAGAGGAA ACAGAATGTG CAAAGACTCA GAATATGGCC TATTTAGGGA 6360
ATGGCTACAT ACACCATGAT TAGAGGAGGC CCAGTAAAGG GAAGGGATGG TGAGATGCCT 6420
GCTAGGTTCA CTCACTCACT TTTATTTATT TATTTATTTT TTTGACAGTC TCTCTGTCGC 6480
CCAGGGCTGGA GTGCAGTGGT GTGATCTTGG GTCACTGCAA CTTCCGCCTC CCGGGTTCAA 6540
GGGATTCTCC TGCCTCAGCT TCCTGAGTAG CTGGGGTTAC AGGTGTGTGC CACCATGCC 6600

AGCTAATTT TTTTGTTATT TTTAGTAGAC AGGGTTTCAC CATGTTGGTC AGGCTGGTCT 6660
CAAACCTCTG GCCTCAAGTG ATCCGCCTGA CTCAGCCTAC CAAAGTGCTG ATTACAAGTG 6720
TGAGGCCACCG TGCCCAGCCA CACTCACTGA TTCTTTAATG CCAGGCCACAC AGCACAAAGT 6780
TCAGAGAAAT GCCTCCATCA TAGCATGTCA ATATGTTCAT ACTCTTAGGT TCATGATGTT 6840
CTTAACATTA GGTTCATATAAG CAAAATAAGA AAAAAGAATA ATAATAAAAA GAAGTGGCAT 6900
GTCAGGACCT CACCTGAAAA GCCAAACACA GAATCATGAA GGTGAATGCA GAGGTGACAC 6960
CAACACAAAG GTGTATATAT GGTTCTCTGT GGGGAGTATG TACGGAGGCA GCAGTGAGTG 7020
AGACTGCAAA CGTCAGAAGG GCACGGGTCA CTGAGAGCCT AGTATCCTAG TAAAGTGGGC 7080
TCTCTCCCTC TCTCTCCAGC TTGTCAATTGA AAACCAGTCC ACCAAGCTTG TTGGTTCGCA 7140
CAGCAAGAGT ACATAGAGTT TGAAATAATA CATAGGATT TAAGAGGGAG AACTGTCTC 7200
TAAAAAAACAAACACAGCA ACAACAAAAA GCAACAAACCA TTACAATTTC ATGTTCCCTC 7260
AGCATTCTCA GAGCTGAGGA ATGGGAGAGG ACTATGGGAA CCCCTTCAT GTTCCGGCCT 7320
TCAGCCATGG CCCTGGATAC ATGCACTCAT CTGTCTTACA ATGTCATTCC CCCAGGAGGG 7380
CCCGGAAGAA AACAGTGCCT ATGAGCAGTT GCTGTCTCGC TTGGAAGAAA TCGCTGAGGA 7440
AGGTCAAGTTT GTTGGTCTGG CCACTAATCT CTGTGGCCTA GTTCATAAAG AATCACCCCTT 7500
TGGAGCTTCA GGTCTGAGGC TGGAGATGGG CTCCCTCCAG TGCAGGAGGG ATTGAAGCAT 7560
GAGCCAGCGC TCATCTTGAT AATAACCATG AAGCTGACAG ACACAGTTAC CCGCAAACGG 7620
CTGCCTACAG ATTGAAAACC AAGCAAAAC CGCCGGGCAC GGTGGCTCAC GCCTGTAATC 7680
CCAGCACTTT GGGAGGCCAA GGCAGGTGGA TCACGGAGTC AAGAGATCAA GACCATCCTG 7740
GCCAACATGG TGAAACCCCA TCTCTACTAA AAATACGAAA AAATAGCCAG GCGTGGTGGC 7800
GGGTGCCTGT AATCCCAGCT ACTCGGGAGG CTGAGGCAGG AGAATGGCAT GAACCCGGGA 7860
GGCAGAAAGTT GCAGTGAGCC GAGATCGTGC CACTGCACTC CAGCCTGGC AACAGAGCGA 7920
GACTCTTGTGTC TCAAAAAAAA AAAAAAAA GAAAACCAAG CAAAAACCAA AATGAGACAA 7980

AAAAAAACAAG ACCAAAAAAAT GGTGTTGGA AATTGTCAAG GTCAAGTCTG GAGAGCTAAA	8040
CTTTTTCTGA GAACTGTTA TCTTTAATAA GCATCAAATA TTTAACTTT GTAAATACTT	8100
TTGTTGGAAA TCGTTCTCTT CTTAGTCACT CTTGGGTCAAT TTTAAATCTC ACTTACTCTA	8160
CTAGACCTTT TAGGTTCTG CTAGACTAGG TAGAACTCTG CCTTGCGATT TCTTGTGTCT	8220
GTTTTGTATA GTTATCAATA TTCATATTAA TTTACAAGTT ATTCAGATCA TTTTTTCTTT	8280
TCTTTTTTTT TTTTTTTT TTTTTTACAT CTTTAGTACA GACAGGGTTT CACCATATTG	8340
GCCAGGCTGC TCTCAAACTC CTGACCTTGT GATCCACCAG CCTCGGCCTC CCAAAGTGCT	8400
GGGATTTCATT TTTTCTTTT AATTTGCTCT GGGCTTAAAC TTGTGGCCCA GCACTTTATG	8460
ATGGTACACA GAGTTAAGAG TGTAGACTCA GACGGTCTTT CTTCTTCCT TCTCTTCCTT	8520
CCTCCCTTCC CTCCCACCTT CCCTTCTCTC CTTCCCTTCT TTCTTCCTCT CTTGCTTCCT	8580
CAGGCCTCTT CCAGTTGCTC CAAAGCCCTG TACTTTTTT TGAGTTAACG TCTTATGGGA	8640
AGGGCCTGCA CTTAGTGAAG AAGTGGTCTC AGAGTTGAGT TACCTTGGCT TCTGGGAGGT	8700
GAAACTGTAT CCCTATAACCC TGAAGCTTTA AGGGGGTGCA ATGTAGATGA GACCCCAACA	8760
TAGATCCTCT TCACAGGCTC AGAGACTCAG GTCCCGAGGAC TGGACATATC TGCACTCCTG	8820
CCCTCTGACT TCAGCCGCTA CTTCCAATAT GAGGGGTCTC TGACTACACC GCCCTGTGCC	8880
CAGGGTGTCA TCTGGACTGT GTTTAACCAAG ACAGTGATGC TGAGTGCTAA GCAGGGGGC	8940
CTGGGGTGTG TGTGGACACA GTGGGTGCGG GGGAAAGAGG ATGTAAGATG AGATGAGAAA	9000
CAGGAGAAGA AAGAAATCAA GGCTGGGCTC TGTGGCTTAC GCCTATAATC CCACCACGTT	9060
GGGAGGCTGA GGTGGGAGAA TGGTTTGAGC CCAGGAGTTC AAGACAAGGC GGGCAACAT	9120
AGTGTGACCC CATCTCTACC AAAAAAACCC CAACAAAACC AAAAATAGCC GGGCATGGTG	9180
GTATGCGGCC TAGTCCCAGC TACTCAAGGA GGCTGAGGTG GGAAGATCGC TTGATTCCAG	9240
GAGTTTGAGA CTGCAGTGAG CTATGATCCC ACCACTGCCT ACCATCTTTA GGATACATTT	9300
ATTTATTTAT AAAAGAAATC AAGAGGCTGG ATGGGAAATA CAGGAGCTGG AGGGTGGAGC	9360

CCTGAGGTGC	TGGTTGTGAG	CTGGCCTGGG	ACCCTTGT	TTC	CCTGTCATGC	CATGAACCCA	9420
CCCACACTGT	CCACTGACCT	CCCTAGCTCC	ACACCCCTCTC	TGACACCC	TGGGGACCTG		9480
GTGACTCTCG	GCTACAGCTG	AACTTCCGAG	CGACGCAGCC	TTTGAATGGG	CGAGTGATTG		9540
AGGCCTCCTT	CCCTGCTGGA	GTGGACAGCA	GTCCTCGGGC	TGCTGAGCCA	GGTACAGCTT		9600
TGTCTGGTTT	CCCCCCAGCC	AGTAGTCCCT	TATCCTCCCA	TGTGTGTGCC	AGTGTCTGTC		9660
ATTGGTGGTC	ACAGCCCCGCC	TCTCACATCT	CCTTTTTCTC	TCCAGTCCAG	CTGAATTCC		9720
GCCTGGCTGC	TGGTGAGTCT	GCCCCTCCTC	TTGGTCCTGA	TGCCAGGAGA	CTCCTCAGCA		9780
CCATTCA	CCAGGGCTGC	TCAGGACCGC	CTCTGCTCCC	TCTCCTTTTC	TGCAGAACAG		9840
ACCCCAACCC	CAATATTAGA	GAGGCAGATC	ATGGTGGGGA	TTCCCCCATT	GTCCCCAGAG		9900
GCTAATTGAT	TAGAATGAAG	CTTGAGAAAT	CTCCCAGCAT	CCCTCTCGCA	AAAGAATCCC		9960
CCCCCCTTTT	TTTAAAGATA	GGGTCTCACT	CTGTTGCC	CAGGCTGGG	TGTTGTGGCA		10020
CGATCATAGC	TCACTGCAGC	CTCGAACTCC	TAGGCTCAGG	CAATCCTTTC	ACCTTAGCTT		10080
CTCAAAGCAC	TGGGACTGTA	GGCATGAGCC	ACTGTGCCTG	GCCCCAAACG	GCCCTTTAC		10140
TTGGCTTTA	GGAAGCAAAA	ACGGTGCTTA	TCTTACCCCT	TCTCGTGTAT	CCACCCCTCAT		10200
CCCTGGCTG	GCCTCTCTG	GAGACTGAGG	CACTATGGG	CTGCCTGAGA	ACTCGGGCA		10260
GGGGTGGTGG	AGTGCACTGA	GGCAGGTGTT	GAGGAACCTCT	GCAGACCC	CTTCCTTCCC		10320
AAAGCAGCCC	TCTCTGCTCT	CCATCGCAGG	TGACATCCTA	GCCCTGGTTT	TTGGCCTCCT		10380
TTTGCTGTC	ACCAGCGTCG	CGTTCTTGT	GCAGATGAGA	AGGCAGCACA	GGTATTACAC		10440
TGACCC	TTCAGGCACA	AGCTTCCCCC	ACCCTTGTGG	AGTCAC	TGCAAAGCGC		10500
ATGCAAATGA	GCTGCTCCTG	GGCCAGTTTT	CTGATTAGCC	TTTCTGTTG	TGTACACACA		10560
GAAGGGGAAC	CAAAGGGGT	GTGAGCTACC	GCCCAGCAGA	GGTAGCCGAG	ACTGGAGCCT		10620
AGAGGCTGGA	TCTTGGAGAA	TGTGAGAAGC	CAGCCAGAGG	CATCTGAGGG	GGAGCCGGTA		10680
ACTGTCCTGT	CCTGCTCATT	ATGCCACTTC	CTTTAACTG	CCAAGAAATT	TTTTAAAATA		10740

AATATTTATA ATAAAATATG TGTTAGTCAC CTTTGTCCC CAAATCAGAA GGAGGTATT 10800
GAATTCCTA TTACTGTTAT TAGCACCAAT TTAGTGGTAA TGCATTTATT CTATTACAGT 10860
TCGGCCTCCT TCCACACATC ACTCCAATGT GTTGCTCC 10898

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: Signal peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Ala Pro Leu Cys Pro Ser Pro Trp Leu Pro Leu Leu Ile Pro Ala
1 5 10 15

Pro Ala Pro Gly Leu Thr Val Gln Leu Leu Leu Ser Leu Leu Leu
20 25 30

Met Pro Val His Pro
35

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TGGGGTTCTT GAGGATCTCC AGGAG

25

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CTCTAACTTC AGGGAGCCCT CTTCTT

26

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "primer"

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (D) OTHER INFORMATION: N stands for inosine

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CUACUACUAC UAGGCCACGC GTCGACTAGT ACGGGNNGGG NNGGGNNG

48

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Glu Glu Asp Leu Pro Ser
1 5

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:

- (A) NAME/KEY: Peptide
- (B) LOCATION: 55..60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Gly Glu Asp Asp Pro Leu
1 5

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Asn Asn Ala His Arg Asp Lys Glu Gly Asp Asp Gln Ser His Trp Arg
1 5 10 15

Tyr Gly Gly Asp Pro
20

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 16 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (v) FRAGMENT TYPE: internal

- (ix) FEATURE:
 - (A) NAME/KEY: Peptide
 - (B) LOCATION: 36..51

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

His Pro Gln Arg Leu Pro Arg Met Gln Glu Asp Ser Pro Leu Gly Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Glu Glu Asp Ser Pro Arg Glu Glu Asp Pro Pro Gly Glu Glu Asp Leu
1 5 10 15

Pro Gly Glu Glu Asp Leu Pro Gly
20

(2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (v) FRAGMENT TYPE: internal
- (ix) FEATURE:
 - (A) NAME/KEY: Peptide
 - (B) LOCATION: 279..291

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Leu Glu Glu Gly Pro Glu Glu Asn Ser Ala Tyr Glu Gln
1 5 10

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Arg Arg Gln His Arg Arg Gly Thr Lys Gly Gly Val Ser Tyr Arg
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GTCGCTAGCT CCATGGGTCA TATGCAGAGG TTGCCCGGA TGCAG

45

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GAAGATCTCT TACTCGAGCA TTCTCCAAGA TCCAGCCTCT AGG

43

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: DNA (genomic)

(A) DESCRIPTION: AP-2 transcription factor

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TCCCCCACCC

10

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: DNA (genomic)

(A) DESCRIPTION: initiator (Inr) element

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

CCACCCCCAT

10

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: DNA (genomic)

- (A) DESCRIPTION: p53 binding site

(x) PUBLICATION INFORMATION:

- (A) AUTHORS: El Deiry et al.
- (B) TITLE: "Human genomic DNA sequences define a consensus binding site for p53"
- (C) JOURNAL: Nature Genetics
- (D) VOLUME: 1
- (F) PAGES: 44-49
- (G) DATE: 1992

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AAGCTAGTCC

10

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Leu Glu His His His His His
1 5

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: Initiator consensus sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

YYYCAYYYYYY

10

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: p53 binding site

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(x) PUBLICATION INFORMATION:

- (A) AUTHORS: El Deiry et al.
- (B) TITLE: "Human genomic DNA sequences define a consensus binding site for p53"
- (C) JOURNAL: Nature Genetics
- (D) VOLUME: 1
- (F) PAGES: 44-49
- (G) DATE: 1992

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

AGGCTTGCTC

10

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Ser Pro Xaa Xaa
1

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Thr Pro Xaa Xaa
1

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 540 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: Proposed MN promoter

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CTTGCTTTTC ATTCAAGCTC AAGTTTGTCT CCCACATACC CATTACTTAA CTCACCCCTCG	60
GGCTCCCCTA GCAGCCTGCC CTACCTCTTT ACCTGCTTCC TGGTGGAGTC AGGGATGTAT	120
ACATGAGCTG CTTTCCCTCT CAGCCAGAGG ACATGGGGGG CCCCAGCTCC CCTGCCTTTC	180
CCCTTCTGTG CCTGGAGCTG GGAAGCAGGC CAGGGTTAGC TGAGGCTGGC TGGCAAGCAG	240
CTGGGTGGTG CCAGGGAGAG CCTGCATAGT GCCAGGTGGT GCCTTGGGTT CCAAGCTAGT	300
CCATGGCCCC GATAACCTTC TGCCTGTGCA CACACCTGCC CCTCACTCCA CCCCCATCCT	360
AGCTTGTTGA TGGGGGAGAG GGCACAGGGC CAGACAAACC TGTGAGACTT TGGCTCCATC	420
TCTGCAAAAG GGCGCTCTGT GAGTCAGCCT GCTCCCTCC AGGCTTGCTC CTCCCCCACC	480
CAGCTCTCGT TTCCAATGCA CGTACAGCCC GTACACACCG TGTGCTGGGA CACCCCACAG	540

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 415 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 1st MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

ACAGTCAGCC GCATGGCTCC CCTGTGCCCG AGCCCCCTGGC TCCCTCTGTT GATCCCGGCC 60
CCTGCTCCAG GCCTCACTGT GCAACTGCTG CTGTCACTGC TGCTTCTGGT GCCTGTCCAT 120
CCCCAGAGGT TGCCCCGGAT GCAGGAGGAT TCCCCCTTGG GAGGAGGCTC TTCTGGGAA 180
GATGACCCAC TGGGCGAGGA GGATCTGCC AGTGAAGAGG ATTCAACCCAG AGAGGAGGAT 240
CCACCCGGAG AGGAGGATCT ACCTGGAGAG GAGGATCTAC CTGGAGAGGA GGATCTACCT 300
GAAGTTAACGC CTAAATCAGA AGAAGAGGGC TCCCTGAAGT TAGAGGATCT ACCTACTGTT 360
GAGGCTCCTG GAGATCCTCA AGAACCCAG AATAATGCC ACAGGGACAA AGAAG 415

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 2nd MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GGGATGACCA GAGTCATTGG CGCTATGGAG 30

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 171 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 3rd MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

GCGACCCGCC CTGGCCCCGG GTGTCCCCAG CCTGCGCGGG CCGCTTCCAG TCCCCGGTGG 60
ATATCCGCC CCAGCTCGCC GCCTTCTGCC CGGCCCTGCG CCCCTGGAA CTCCTGGGCT 120
TCCAGCTCCC GCCGCTCCA GAACTGCGCC TGCGCAACAA TGGCCACAGT G 171

(2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 143 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 4th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

TGCAACTGAC CCTGCCTCCT GGGCTAGAGA TGGCTCTGGG TCCCCGGCGGG GAGTACCGGG 60
CTCTGCAGCT GCATCTGCAC TGGGGGGCTG CAGGTCGTCC GGGCTCGGAG CACACTGTGG 120

AAGGCCACCG TTTCCCTGCC GAG

143

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 93 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 5th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

ATCCACGTGG TTCACCTCAG CACCGGCTTT GCCAGAGTTG ACGAGGCCTT GGGCGCCCG

60

GGAGGGCTGG CCGTGTTGGC CGCCTTTCTG GAG

93

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 67 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 6th MN exon

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

GAGGGCCCGG AAGAAAACAG TGCCTATGAG CAGTTGCTGT CTCGCTTGGA AGAAATCGCT

60

GAGGAAG

67

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 158 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 7th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GCTCAGAGAC TCAGGTCCA GGACTGGACA TATCTGCACT CCTGCCCTCT GACTTCAGCC	60
GCTACTTCCA ATATGAGGGG TCTCTGACTA CACCGCCCTG TGCCCAGGGT GTCATCTGGA	120
CTGTGTTAA CCAGACAGTG ATGCTGAGTG CTAAGCAG	158

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 145 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 8th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CTCCACACCC TCTCTGACAC CCTGTGGGGA CCTGGTGACT CTCGGCTACA GCTGAACCTC	60
CGAGCGACGC AGCCTTGAA TGGGCGAGTG ATTGAGGCCT CCTTCCCTGC TGGAGTGGAC	120
AGCAGTCCTC GGGCTGCTGA GCCAG	145

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 9th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

TCCAGCTGAA TTCCTGCCTG GCTGCTG 27

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 82 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 10th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GTGACATCCT AGCCCTGGTT TTTGGCCTCC TTTTGCTGT CACCAGCGTC GCGTTCCTTG 60

TGCAGATGAG AAGGCAGCAC AG 82

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 11th MN exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

AAGGGGAACC AAAGGGGGTG TGAGCTACCG CCCAGCAGAG GTAGCCGAGA CTGGAGCCTA 60

GAGGCTGGAT CTTGGAGAAC GTGAGAAAGCC AGCCAGAGGC ATCTGAGGGG GAGCCGGTAA 120

CTGTCCTGTC CTGCTCATTA TGCCACTTCC TTTAACTGC CAAGAAATT TTTAAAATAA 180

ATATTATATAA T 191

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1174 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 1st MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

GTAAGTGGTC ATCAATCTCC AAATCCAGGT TCCAGGAGGT TCATGACTCC CCTCCCATA	60
CCCAGCCTAG GCTCTGTTCA CTCAGGGAAG GAGGGGAGAC TGTACTCCCC ACAGAAGCCC	120
TTCCAGAGGT CCCATACCAA TATCCCCATC CCCACTCTCG GAGGTAGAAA GGGACAGATG	180
TGGAGAGAAA ATAAAAAGGG TGCAAAAGGA GAGAGGTGAG CTGGATGAGA TGGGAGAGAA	240
GGGGGAGGCT GGAGAACAGA AAGGGATGAG AACTGCAGAT GAGAGAAAAA ATGTGCAGAC	300
AGAGGAAAAA AATAGGTGGA GAAGGGAGGT CAGAGAGTTT GAGGGGAAGA GAAAAGGAAA	360
GCTTGGGAGG TGAAGTGGGT ACCAGAGACA AGCAAGAAGA GCTGGTAGAA GTCATCTCAT	420
CTTAGGCTAC AATGAGGAAT TGAGACCTAG GAAGAACGGGA CACAGCAGGT AGAGAACGT	480
GGCTTCTTGA CTCCAAGCC AGGAATTGG GGAAAGGGGT TGGAGACCAT ACAAGGCAGA	540
GGGATGAGTG GGGAGAACAGA AGAAGGGAGA AAGGAAAGAT GGTGTACTCA CTCATTGGG	600
ACTCAGGACT GAAAGTGCCA CTCACTTTT TTTTTTTTT TTTTGAGACA AACCTTCACT	660
TTTGTGCCC AGGCTGGAGT GCAATGGCGC GATCTCGGCT CACTGCAACC TCCACCTCCC	720
GGGTTCAAGT GATTCTCCTG CCTCAGCCTC TAGCCAAGTA GCTGCGATTA CAGGCATGCG	780
CCACCACGCC CGGCTAATT TTGTATTTT AGTAGAGACG GGGTTCGCC ATGTTGGTCA	840
GGCTGGTCTC GAACTCCTGA TCTCAGGTGA TCCAACCACC CTGGCCTCCC AAAGTGCTGG	900
GATTATAGGC GTGAGCCACA GCGCCTGGCC TGAAGCAGCC ACTCACTTTT ACAGACCCTA	960
AGACAATGAT TGCAAGCTGG TAGGATTGCT GTTGGCCCA CCCAGCTGCG GTGTTGAGTT	1020
TGGGTGCGGT CTCCTGTGCT TTGCACCTGG CCCGCTTAAG GCATTGTTA CCCGTAATGC	1080
TCCTGTAAGG CATCTGCGTT TGTGACATCG TTTGGTGC CAGGAAGGGA TTGGGGCTCT	1140
AAGCTTGAGC GGTCATCCT TTTCATTAT ACAG	1174

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 193 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 2nd MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

GTGAGACACC CACCCGCTGC ACAGACCCAA TCTGGGAACC CAGCTCTGTG GATCTCCCCT	60
ACAGCCGTCC CTGAACACTG GTCCCCGGCG TCCCACCCGC CGCCCACCGT CCCACCCCT	120
CACCTTTCT ACCCGGGTTC CCTAAGTTCC TGACCTAGGC GTCAGACTTC CTCACTATAAC	180
TCTCCCACCC CAG	193

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 131 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 3rd MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GTGAGGGGGT CTCCCCGCCG AGACTTGGGG ATGGGGCGGG GCGCAGGGAA GGGAAACCGTC	60
GCGCAGTGCC TGCCCAGGGGG TTGGGCTGGC CCTACCGGGC GGGGCCGGCT CACTTGCCTC	120
TCCCTACGCA G	131

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 89 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 4th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

GTGAGCGCGG ACTGGCCGAG AAGGGGCAA GGAGCGGGGC GGACGGGGC CAGAGACGTG	60
GCCCTCTCCT ACCCTCGTGT CCTTTTCAG	89

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1400 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

GTACCAGATC	CTGGACACCC	CCTACTCCCC	GCTTTCCCAT	CCCATGCTCC	TCCCGGACTC	60
TATCGTGGAG	CCAGAGACCC	CATCCCAGCA	AGCTCACTCA	GGCCCCTGGC	TGACAAACTC	120
ATTCACGCAC	TGTTTGTCA	TTAACACCC	ACTGTGAACC	AGGCACCAGC	CCCCAACAAAG	180
GATTCTGAAG	CTGTAGGTCC	TTGCCTCTAA	GGAGCCCACA	GCCAGTGGGG	GAGGCTGACA	240
TGACAGACAC	ATAGGAAGGA	CATAGTAAAG	ATGGTGGTCA	CAGAGGAGGT	GACACTTAAA	300
GCCTTCACTG	GTAGAAAAGA	AAAGGAGGTG	TTCATTGCAG	AGGAAACAGA	ATGTGCAAAG	360
ACTCAGAATA	TGGCCTATT	AGGGAATGGC	TACATACACC	ATGATTAGAG	GAGGCCAGT	420
AAAGGGAAGG	GATGGTGAGA	TGCCTGCTAG	GTTCACTCAC	TCACTTTAT	TTATTATTT	480
ATTTTTTG	CAGTCTCTCT	GTCGCCAGG	CTGGAGTGCA	GTGGTGTGAT	CTTGGGTCAC	540
TGCAACTTCC	GCCTCCCGGG	TTCAAGGGAT	TCTCCTGCCT	CAGCTTCCTG	AGTAGCTGGG	600
GTTACAGGTG	TGTGCCACCA	TGCCAGCTA	ATTTTTTT	GTATTTTAG	TAGACAGGGT	660
TTCACCATGT	TGGTCAGGCT	GGTCTAAAC	TCCTGGCCTC	AAGTGATCCG	CCTGACTCAG	720
CCTACCAAAG	TGCTGATTAC	AAGTGTGAGC	CACCGTGCCC	AGCCACACTC	ACTGATTCTT	780
TAATGCCAGC	CACACAGCAC	AAAGTTCAGA	GAAATGCCTC	CATCATAGCA	TGTCAATATG	840
TTCATACTCT	TAGGTTCATG	ATGTTCTTAA	CATTAGGTT	ATAAGCAAAA	TAAGAAAAAA	900
GAATAATAAA	AAAAAGAAGT	GGCATGTCAG	GACCTCACCT	GAAAAGCCAA	ACACAGAAC	960
ATGAAGGTGA	ATGCAGAGGT	GACACCAACA	CAAAGGTGTA	TATATGGTTT	CCTGTGGGGA	1020

GTATGTACGG AGGCAGCACT GAGTGAGACT GCAAACGTCA GAAGGGCACG GGTCACTGAG	1080
AGCCTAGTAT CCTAGTAAAG TGGGCTCTCT CCCTCTCTCT CCAGCTTGTC ATTGAAAACC	1140
AGTCCACCAA GCTTGTGGT TCGCACAGCA AGAGTACATA GAGTTTGAAA TAATACATAG	1200
GATTTAAGA GGGAGACACT GTCTCTAAAA AAAAAAACAA CAGCAACAAC AAAAAGCAAC	1260
AACCATTACA ATTTTATGTT CCCTCAGCAT TCTCAGAGCT GAGGAATGGG AGAGGACTAT	1320
GGGAACCCCC TTCATGTTCC GCCCTTCAGC CATGCCCTG GATACATGCA CTCATCTGTC	1380
TTACAATGTC ATTCCCCCAG	1400

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1334 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 6th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GTCAGTTGT TGGTCTGGCC ACTAATCTCT GTGGCCTAGT TCATAAAGAA TCACCCTTG	60
GAGCTTCAGG TCTGAGGCTG GAGATGGGCT CCCTCCAGTG CAGGAGGGAT TGAAGCATGA	120
GCCAGCGCTC ATCTTGATAA TAACCATGAA GCTGACAGAC ACAGTTACCC GCAAACGGCT	180
GCCTACAGAT TGAAAACCAA GCAAAACCG CGGGCACGG TGGCTCACGC CTGTAATCCC	240
AGCACTTTGG GAGGCCAAGG CAGGTGGATC ACGAGGTCAA GAGATCAAGA CCATCCTGGC	300

CAACATGGTG AAACCCCATC TCTACTAAAA ATACGAAAAA ATAGCCAGGC GTGGTGGCGG	360
GTGCCTGTAA TCCCAGCTAC TCGGGAGGCT GAGGCAGGAG AATGGCATGA ACCCGGGAGG	420
CAGAAGTTGC AGTGAGCCGA GATCGTGCCA CTGCACTCCA GCCTGGCAA CAGAGCGAGA	480
CTCTTGTCTC AAAAAAAA AAAAAGA AAACCAAGCA AAAACCAAAA TGAGACAAAA	540
AAAACAAGAC CAAAAAATGG TGTTTGGAAA TTGTCAAGGT CAAGTCTGGA GAGCTAACT	600
TTTTCTGAGA ACTGTTTATC TTTAATAAGC ATCAAATATT TTAACTTGT AAATACTTT	660
GTTGGAAATC GTTCTCTCT TAGTCACTCT TGGGTCAATT TAAATCTCAC TTACTCTACT	720
AGACCTTTA GGTTCTGCT AGACTAGGTA GAACTCTGCC TTTGCATTTC TTGTGTCTGT	780
TTTGTATAGT TATCAATATT CATATTATT TACAAGTTAT TCAGATCATT TTTCTTTTC	840
TTTTTTTTT TTTTTTTTT TTTTACATCT TTAGTAGAGA CAGGGTTCA CCATATTGGC	900
CAGGCTGCTC TCAAACCTCCT GACCTTGTGA TCCACCAGCC TCAGCCTCCC AAAGTGCTGG	960
GATTCAATTAA TTCTTTTAA TTGCTCTGG GCTTAAACTT GTGGCCAGC ACTTTATGAT	1020
GGTACACAGA GTTAAGAGTG TAGACTCAGA CGGTCTTCT TCTTCCTTC TCTTCCTCC	1080
TCCCTCCCT CCCACCTTCC CTTCTCTCCT TCCTTTCTTT CTTCCCTCT TGCTTCCTCA	1140
GGCCTCTCC AGTTGCTCCA AAGCCCTGTA CTTTTTTTG AGTTAACGTC TTATGGGAAG	1200
GGCCTGCACT TAGTGAAGAA GTGGTCTCAG AGTTGAGTTA CCTTGGCTTC TGGGAGGTGA	1260
AACTGTATCC CTATACCTG AAGCTTAAG GGGGTGCAAT GTAGATGAGA CCCAACATA	1320
GATCCTCTTC ACAG	1334

(2) INFORMATION FOR SEQ ID NO: 45:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 512 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 7th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GTGGGCCTGG GGTGTGTGTG GACACAGTGG GTGCGGGGGA AAGAGGATGT AAGATGAGAT	60
GAGAAACAGG AGAAGAAAGA AATCAAGGCT GGGCTCTGTG GCTTACGCCT ATAATCCCAC	120
CACGTTGGGA GGCTGAGGTG GGAGAATGGT TTGAGCCCAG GAGTTCAAGA CAAGGCGGGG	180
CAACATAGTG TGACCCCATC TCTACCAAAA AAACCCCAAC AAAACCAAAA ATAGCCGGGC	240
ATGGTGGTAT GCggcctagt CCCAGCTACT CAAGGAGGCT GAGGTGGAA GATCGCTTGA	300
TTCCAGGAGT TTGAGACTGC AGTGAGCTAT GATCCCACCA CTGCCTACCA TCTTTAGGAT	360
ACATTATTT ATTTATAAAA GAAATCAAGA GGCTGGATGG GGAATACAGG AGCTGGAGGG	420
TGGAGCCCTG AGGTGCTGGT TGTGAGCTGG CCTGGGACCC TTGTTTCCTG TCATGCCATG	480
AACCCACCCA CACTGTCCAC TGACCTCCCT AG	512

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 114 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 8th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

GTACAGCTT GTCTGGTTTC CCCCCAGCCA GTAGTCCCTT ATCCTCCCAT GTGTGTGCCA 60

GTGTCTGTCA TTGGTGGTCA CAGCCCGCCT CTCACATCTC CTTTTCTCT CCAG 114

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 617 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 9th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

GTGAGTCTGC CCCTCCTCTT GGTCTGATG CCAGGAGACT CCTCAGCACC ATTCAAGCCCC 60

AGGGCTGCTC AGGACCGCCT CTGCTCCCTC TCCTTTCTG CAGAACAGAC CCCAACCCCA 120

ATATTAGAGA GGCAGATCAT GGTGGGGATT CCCCCATTGT CCCCAGAGGC TAATTGATTA 180

GAATGAAGCT TGAGAAATCT CCCAGCATCC CTCTCGAAA AGAATCCCCC CCCCTTTTTT 240

TAAAGATAAG GTCTCACTCT GTTGCCCCA GGCTGGGGTG TTGTGGCACG ATCATAGCTC 300

ACTGCAGCCT CGAACTCCTA GGCTCAGGCA ATCCTTCAC CTTAGCTTCT CAAAGCACTG 360

GGACTGTAGG CATGAGCCAC TGTGCCTGGC CCCAAACGGC CCTTTTACTT GGCTTTAGG 420

AAGCAAAAC GGTGCTTATC TTACCCCTTC TCGTGTATCC ACCCTCATCC CTTGGCTGGC 480

CTCTTCTGGA GACTGAGGCA CTATGGGGCT GCCTGAGAAC TCGGGGCAGG GGTGGTGGAG 540

TGCAC TGAGG CAGGT GTTGA GGA ACTCTGC AG ACC CCTCT TCCTT CCCAA AGCAG CCCTC	600
TCTGCTCTCC ATCGCAG	617

(2) INFORMATION FOR SEQ ID NO: 48:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 130 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 10th MN intron

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

GTATTACACT GACCCTTTCT TCAGGCACAA GCTTCCCCCA CCCTTGTTGA GTCAC TTCAT	60
GCAAAGCGCA TGCAAATGAG CTGCTCCTGG GCCAGTTTC TGATTAGCCT TTCCCTGTTGT	120
GTACACACAG	130

(2) INFORMATION FOR SEQ ID NO: 49:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1401 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: Spans 3' part of 1st intron to beyond end of 5th exon

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

CAAACTTCA	CTTTTGTG C	CCAGGCTGGA	GTGCAATGGC	GCGATCTCGG	CTCACTGCAA	60
CCTCCACCTC	CCGGGTTCAA	GTGATTCTCC	TGCCTCAGCC	TCTAGCCAAG	TAGCTGCGAT	120
TACAGGCATG	CGCCACCACG	CCC GGCTAAT	TTTTGTATTT	TTAGTAGAGA	CGGGGTTTCG	180
CCATGTTGGT	CAGGCTGGTC	TCGAACTCCT	GATCTCAGGT	GATCCAACCA	CCCTGGCCTC	240
CCAAAGTGCT	GGGATTATAG	GCGTGAGCCA	CAGCGCCTGG	CCTGAAGCAG	CCACTCACTT	300
TTACAGACCC	TAAGACAATG	ATTGCAAGCT	GGTAGGATTG	CTGTTGGCC	CACCCAGCTG	360
CGGTGTTGAG	TTTGGGTGCG	GTCTCCTGTG	CTTGCACCT	GGCCCGCTTA	AGGCATTGT	420
TACCCGTAAT	GCTCCTGTAA	GGCATCTGCG	TTTGTGACAT	CGTTTGTC	GCCAGGAAGG	480
GATTGGGGCT	CTAACGTTGA	GCGGTTCATC	CTTTTCATTT	ATACAGGGGA	TGACCAGAGT	540
CATTGGCGCT	ATGGAGGTGA	GACACCCACC	CGCTGCACAG	ACCCAATCTG	GGAACCCAGC	600
TCTGTGGATC	TCCCCTACAG	CCGTCCTGA	ACACTGGTCC	CGGGCGTCCC	ACCCGCCGCC	660
CACCGTCCA	CCCCCTCACC	TTTCTACCC	GGGTTCCCTA	AGTCCTGAC	CTAGGCGTCA	720
GACTTCCTCA	CTATACTCTC	CCACCCCAGG	CGACCCGCC	TGGCCCCGGG	TGTCCCCAGC	780
CTGCGCGGGC	CGCTTCCAGT	CCCCGGTGGA	TATCCGCCCC	CAGCTCGCCG	CCTTCTGCC	840
GGCCCTGCGC	CCCCTGGAAC	TCCTGGGCTT	CCAGCTCCCG	CCGCTCCAG	AACTGCGCCT	900
GCGCAACAAT	GGCCACAGTG	GTGAGGGGGT	CTCCCCGCCG	AGACTTGGGG	ATGGGGCGGG	960
GCGCAGGGAA	GGGAACCGTC	GCGCAGTGCC	TGCCCCGGGG	TTGGGCTGGC	CCTACCGGGC	1020
GGGGCCGGCT	CACTTGCCTC	TCCCTACGCA	GTGCAACTGA	CCCTGCCTCC	TGGGCTAGAG	1080
ATGGCTCTGG	GTCCCAGGCG	GGAGTACCGG	GCTCTGCAGC	TGCATCTGCA	CTGGGGGGCT	1140

GCAGGGTCGTC CGGGCTCGGA GCACACTGTG GAAGGCCACC GTTCCCTGC CGAGGTGAGC	1200
GC GGACTGGC CGAGAAGGGG CAAAGGAGCG GGGCGGACGG GGGCCAGAGA CGTGGCCCTC	1260
TCCTACCCCTC GTGTCCCTTT CAGATCCACG TGTTCACCT CAGCACCGCC TTTGCCAGAG	1320
TTGACGGAGGC CTTGGGGCGC CCGGGAGGCC TGGCCGTGTT GGCCGCCTT CTGGAGGTAC	1380
CAGATCCTGG ACACCCCCCTA C	1401

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 98 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(A) DESCRIPTION: Region of homology to collagen alpha 1 chain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

Gln	Arg	Leu	Pro	Arg	Met	Gln	Glu	Asp	Ser	Pro	Leu	Gly	Gly	Ser
1					5					10			15	

Ser	Gly	Glu	Asp	Asp	Pro	Leu	Gly	Glu	Glu	Asp	Leu	Pro	Ser	Glu	Glu
					20			25				30			

Asp	Ser	Pro	Arg	Glu	Glu	Asp	Pro	Pro	Gly	Glu	Glu	Asp	Leu	Pro	Gly
						35		40				45			

Glu	Glu	Asp	Leu	Pro	Gly	Glu	Glu	Asp	Leu	Pro	Glu	Val	Lys	Pro	Lys
						50		55			60				

Ser	Glu	Glu	Gly	Ser	Leu	Lys	Leu	Glu	Asp	Leu	Pro	Thr	Val	Glu
					65		70		75			80		

Ala	Pro	Gly	Asp	Pro	Gln	Glu	Pro	Gln	Asn	Asn	Ala	His	Arg	Asp	Lys
						85			90			95			

Glu Gly

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 256 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(A) DESCRIPTION: carbonic anhydrase domain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

Asp Asp Gln Ser His Trp Arg Tyr Gly Gly Asp Pro Pro Trp Pro Arg
1 5 10 15

Val Ser Pro Ala Cys Ala Gly Arg Phe Gln Ser Pro Val Asp Ile Arg
20 25 30

Pro Gln Leu Ala Ala Phe Cys Pro Ala Leu Arg Pro Leu Glu Leu Leu
35 40 45

Gly Phe Gln Leu Pro Pro Leu Pro Glu Leu Arg Leu Arg Asn Asn Gly
50 55 60

His Ser Val Gln Leu Thr Leu Pro Pro Gly Leu Glu Met Ala Leu Gly
65 70 75 80

Pro Gly Arg Glu Tyr Arg Ala Leu Gln Leu His Leu His Trp Gly Ala
85 90 95

Ala Gly Arg Pro Gly Ser Glu His Thr Val Glu Gly His Arg Phe Pro
100 105 110

Ala Glu Ile His Val Val His Leu Ser Thr Ala Phe Ala Arg Val Asp
115 120 125

Glu Ala Leu Gly Arg Pro Gly Gly Leu Ala Val Leu Ala Ala Phe Leu
130 135 140

Glu Glu Gly Pro Glu Glu Asn Ser Ala Tyr Glu Gln Leu Leu Ser Arg
145 150 155 160

Leu Glu Glu Ile Ala Glu Glu Gly Ser Glu Thr Gln Val Pro Gly Leu
165 170 175

Asp Ile Ser Ala Leu Leu Pro Ser Asp Phe Ser Arg Tyr Phe Gln Tyr
180 185 190

Glu Gly Ser Leu Thr Thr Pro Pro Cys Ala Gln Gly Val Ile Trp Thr
195 200 205

Val Phe Asn Gln Thr Val Met Leu Ser Ala Lys Gln Leu His Thr Leu
210 215 220

Ser Asp Thr Leu Trp Gly Pro Gly Asp Ser Arg Leu Gln Leu Asn Phe
225 230 235 240

Arg Ala Thr Gln Pro Leu Asn Gly Arg Val Ile Glu Ala Ser Phe Pro
245 250 255

(2) INFORMATION FOR SEQ ID NO: 52:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(A) DESCRIPTION: transmembrane region

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

Asp Ile Leu Ala Leu Val Phe Gly Leu Leu Phe Ala Val Thr Ser Val
1 5 10 15

Ala Phe Leu Val
20

(2) INFORMATION FOR SEQ ID NO: 53:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 amino acids
 - (B) TYPE: amino acid

- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (A) DESCRIPTION: intracellular C-terminus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

Met	Arg	Arg	Gln	His	Arg	Arg	Gly	Thr	Lys	Gly	Gly	Val	Ser	Tyr	Arg
1				5					10					15	
Pro	Ala	Glu	Val	Ala	Glu	Thr	Gly	Ala							
			20					25							

(2) INFORMATION FOR SEQ ID NO: 54:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 170 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

Arg	Ala	Leu	Gln	Leu	His	Leu	His	Trp	Gly	Ala	Ala	Gly	Arg	Pro	Gly
1				5				10					15		
Ser	Glu	His	Thr	Val	Glu	Gly	His	Arg	Phe	Pro	Ala	Glu	Ile	His	Val
			20				25					30			
Val	His	Leu	Ser	Thr	Ala	Phe	Ala	Arg	Val	Asp	Glu	Ala	Leu	Gly	Arg
	35					40						45			
Pro	Gly	Gly	Leu	Ala	Val	Leu	Ala	Ala	Phe	Leu	Glu	Gly	Pro	Glu	
			50			55					60				
Glu	Asn	Ser	Ala	Tyr	Glu	Gln	Leu	Leu	Ser	Arg	Leu	Glu	Glu	Ile	Ala
	65				70				75				80		
Glu	Glu	Gly	Ser	Glu	Thr	Gln	Val	Pro	Gly	Leu	Asp	Ile	Ser	Ala	Leu

85	90	95
Leu Pro Ser Asp Phe Ser Arg Tyr Phe Gln Tyr Glu Gly Ser Leu Thr		
100	105	110
Thr Pro Pro Cys Ala Gln Gly Val Ile Trp Thr Val Phe Asn Gln Thr		
115	120	125
Val Met Leu Ser Ala Lys Gln Leu His Thr Leu Ser Asp Thr Leu Trp		
130	135	140
Gly Pro Gly Asp Ser Arg Leu Gln Leu Asn Phe Arg Ala Thr Gln Pro		
145	150	155
Leu Asn Gly Arg Val Ile Glu Ala Ser Phe		
165	170	

(2) INFORMATION FOR SEQ ID NO: 55:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 470 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: RNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

CAUGGCCCG AUAACCUUCU GCCUGUGCAC ACACCUGCCC CUCACUCCAC CCCCAUCCUA	60
GUUUUGGUAU GGGGGAGAGG GCACAGGGCC AGACAAACCU GUGAGACUUU GGCUCCAUCU	120
CUGCAAAAGG GCGCUCUGUG AGUCAGCCUG CUCCCCUCCA GGCUUGCUCUCC UCCCCCACCC	180
AGCUCUCGUU UCCAAUGCAC GUACAGCCCG UACACACCGU GUGCUGGGAC ACCCCACAGU	240
CAGCCGCAUG GCUCUCCCUGU GCCCCAGCCC CUGGCUCCCU CUGUUGAUCC CGGCCCCUGC	300
UCCAGGCCUC ACUGUGCAAC UGCUGCUGUC ACUGGCUGUU CUGGUGCCUG UCCAUCCCCA	360
GAGGUUGCCC CGGAUGCAGG AGGAUUCCCC CUUGGGAGGA GGCUCUUCUG GGGAAAGAUGA	420

CCCACUGGGC GAGGAGGAUC UGCCAGUGA AGAGGAUUC CCCAGAGAGG

470

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

This sequence is intentionally skipped.

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

This sequence is intentionally skipped.

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 904 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

GCTGGTCTCG AACTCCTGGA CTCAAGCAAT CCACCCACCT CAGCCTCCCA AAATGAGGGA

60

CCGTGTCTTA TTCATTCCA TGTCCCTAGT CCATAGCCCA GTGCTGGACC TATGGTAGTA

120

CTAAATAAAAT ATTTGTTGAA TGCAATAGTA AATAGCATT CAGGGAGCAA GAACTAGATT

180

AACAAAGGTG GTAAAAGGTT TGGAGAAAAA AATAATAGTT TAATTGGCT AGAGTATGAG

240

GGAGAGTAGT AGGAGACAAG ATGGAAAGGT CTCTGGGCA AGGTTTGAA GGAAGTTGGA	3 00
AGTCAGAAGT ACACAATGTG CATATCGTGG CAGGCAGTGG GGAGCCAATG AAGGCTTTG	3 60
AGCAGGAGAG TAATGTGTTG AAAAATAAAT ATAGGTTAAA CCTATCAGAG CCCCTCTGAC	4 20
ACATACACTT GCTTTTCATT CAAGCTCAAG TTTGTCTCCC ACATACCCAT TACTTAAC	4 80
ACCCTCGGGC TCCCCTAGCA GCCTGCCCTA CCTCTTTACC TGCTTCCTGG TGGAGTCAGG	5 40
GATGTATACA TGAGCTGCTT TCCCTCTCAG CCAGAGGACA TGGGGGGCCC CAGCTCCCCT	6 00
GCCTTTCCCC TTCTGTGCCT GGAGCTGGGA AGCAGGCCAG GGTTAGCTGA GGCTGGCTGG	6 60
CAAGCAGCTG GGTGGTGCCA GGGAGAGCCT GCATAGTGCC AGGTGGTGCC TTGGGTTCCA	7 20
AGCTAGTCCA TGGCCCCGAT AACCTTCTGC CTGTGCACAC ACCTGCCCT CACTCCACCC	7 80
CCATCCTAGC TTTGGTATGG GGGAGAGGGC ACAGGGCCAG ACAAACCTGT GAGACTTTGG	8 40
CTCCATCTCT GCAAAAGGGC GCTCTGTGAG TCAGCCTGCT CCCCTCCAGG CTTGCTCCTC	9 00
CCCC	9 04

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 292 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

TTTTTTTGAG ACGGAGTCTT GCATCTGTCA TGCCCAGGCT GGAGTAGGAG TGGTGCCATC	6 0
TCGGCTCACT GCAAGCTCCA CCTCCCGAGT TCACGCCATT TTCCTGCCTC AGCCTCCCGA	12 0

GTAGCTGGGA CTACAGGCGC CCGCCACCAT GCCCGGCTAA TTTTTGTAT TTTTGGTAGA	180
GACGGGGTTT CACCGTGTAA GCCAGAATGG TCTCGATCTC CTGACTTCGT GATCCACCCG	240
CCTCGGCCTC CCAAAGTTCT GGGATTACAG GTGTGAGCCA CCGCACCTGG CC	292

(2) INFORMATION FOR SEQ ID NO: 60:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 262 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

TTCTTTTTG AGACAGGGTC TTGCTCTGTC ACCCAGGCCA GAGTGCAATG GTACAGTCTC	60
AGCTCACTGC AGCCTCAACC GCCTCGGCTC AAACCATCAT CCCATTCAG CCTCCTGAGT	120
AGCTGGGACT ACAGGCACAT GCCATTACAC CTGGCTAATT TTTTTGTATT TCTAGTAGAG	180
ACAGGGTTTGC GCCATGTTGC CCGGGCTGGT CTCGAACCTCC TGGACTCAAG CAATCCACCC	240
ACCTCAGCCT CCCAAAATGA GG	262

(2) INFORMATION FOR SEQ ID NO: 61:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TTTTTTTTTG AGACAAACTT TCACTTTGT TGCCCAGGCT GGAGTGCAAT GGCGCGATCT	60
CGGCTCACTG CAACCTCCAC CTCCCGGGTT CAAGTGATTG TCCTGCCTCA GCCTCTAGCC	120
AAGTAGCTGC GATTACAGGC ATGCGCCACC ACGCCCGGCT AATTTTGTA TTTTAGTAG	180
AGACGGGGTT TCGCCATGTT GGTCAGGCTG GTCTCGAACT CCTGATCTCA GGTGATCCAA	240
CCACCCTGGC CTCCCAAAGT GCTGGGATTA TAGGCGTGAG CCACAGCGCC TGGC	294

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 276 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

TGACAGTCTC TCTGTCGCC AGGCTGGAGT GCAGTGGTGT GATCTGGGT CACTGCAACT	60
TCCGCCTCCC GGGTTCAAGG GATTCTCCTG CCTCAGCTTC CTGAGTAGCT GGGGTTACAG	120
GTGTGTGCCA CCATGCCAG CTAATTTTT TTTGTATTT TAGTAGACAG GGTTTCACCA	180
TGTTGGTCAG GCTGGTCTCA AACTCCTGGC CTCAAGTGAT CCGCCTGACT CAGCCTACCA	240
AAGTGCTGAT TACAAGTGTG AGCCACCGTG CCCAGC	276

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 289 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

CGCCGGGCAC GGTGGCTCAC GCCTGTAATC CCAGCACTTT GGGAGGCCAA GGCAGGTGGA	60
TCACGAGGTC AAGAGATCAA GACCATCCTG GCCAACATGG TGAAACCCCA TCTCTACTAA	120
AAATACGAAA AAATAGCCAG GCGTGGTGGC GGGTGCCTGT AATCCCAGCT ACTCGGGAGG	180
CTGAGGCAGG AGAATGGCAT GAACCCGGGA GGCAGAAGTT GCAGTGAGCC GAGATCGTGC	240
CACTGCACTC CAGCCTGGGC AACAGAGCGA GACTCTTGTGTC TCAAAAAAAA	289

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 298 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

AGGCTGGGCT	CTGTGGCTTA	CGCCTATAAT	CCCACCACGT	TGGGAGGCTG	AGGTGGGAGA	60
ATGGTTGAG	CCCAGGAGTT	CAAGACAAGG	CGGGGCAACA	TAGTGTGACC	CCATCTCTAC	120
CAAAAAAAACC	CCAACAAAAC	CAAAAATAGC	CGGGCATGGT	GGTATGCGGC	CTAGTCCCAG	180
CTACTCAAGG	AGGCTGAGGT	GGGAAGATCG	CTTGATTCCA	GGAGTTTGAG	ACTGCAGTGA	240
GCTATGATCC	CACCACTGCC	TACCATCTT	AGGATACATT	TATTTATTTA	TAAAAGAA	298

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 105 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

TTTTTACAT	CTTTAGTAGA	GACAGGGTTT	CACCATATTG	GCCAGGCTGC	TCTCAAACTC	60
CTGACCTTGT	GATCCACCAAG	CCTCGGCCTC	CCAAAGTGCT	GGGAT		105

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 83 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

CCTCGAACTC CTAGGCTCAG GCAATCCTTT CACCTTAGCT TCTCAAAGCA CTGGGACTGT 60
AGGCATGAGC CACTGTGCCT GGC 83

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

AGAAGGTAAG T 11

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

TGGAGGTGAG A 11

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

CAGTCGTGAG G

11

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

CCGAGGGTGAG C

11

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

TGGAGGTACC A

11

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

GGAAGGTCAG T

11

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

AGCAGGTGGG C

11

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

GCCAGGTACA G

11

(2) INFORMATION FOR SEQ ID NO: 75:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

TGCTGGTGAG T

11

(2) INFORMATION FOR SEQ ID NO: 76:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 5' donor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

ATACAGGGGAT

11

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

ATACAGGGGA T

11

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

CCCCAGGCGA C

11

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

ACGCAGTGCA A

11

(2) INFORMATION FOR SEQ ID NO: 80:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

TTTCAGATCC A

11

(2) INFORMATION FOR SEQ ID NO: 81:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

CCCCAGGAGG G

11

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

TCACAGGCTC A

11

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

CCCTAGCTCC A

11

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

CTCCAGTCCA G

11

(2) INFORMATION FOR SEQ ID NO: 85:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

TCGCAGGTGA CA

12

(2) INFORMATION FOR SEQ ID NO: 86:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (A) DESCRIPTION: 3' acceptor consensus splice sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

ACACAGAAGG G